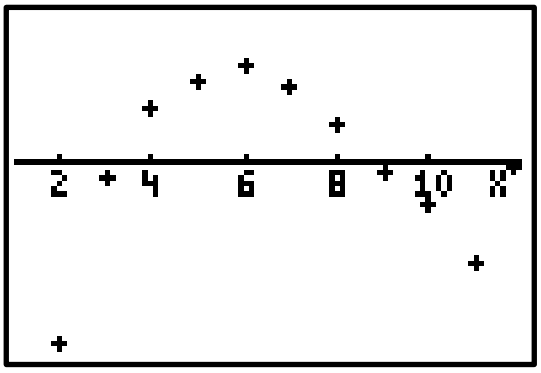
Test: Bivariate Data: Regression Analysis and Two-Way Tables
Instructor's Version

This space intentionally left blank.

Answers: Free Response: 3 questions, 52 points, 30 minutes

1. The following graph is a residual plot derived from a scatterplot that predicted the height of girls from their ages (in years).



A. Does the residual plot indicate that a line is a good model for predicting the height of girls from their ages? Explain. (10 points)

Answer
The residual plot indicates that a line is not a good model. The pattern should be more-or-less random about the *x*-axis. The extreme curvature pattern indicates a non-linear pattern in the original data.

© Copyright 1999-2009 Apex Learning, Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning, Inc.

10 points for this part – holistic based on the degree to which they adequately describe the failure of the plot to justify a linear model.

B. If you were using the regression line to predict the *y*-value for x = 5.5 years, would your prediction most likely be too large or too small? Explain your reasoning. (10 points)

Answer
It's reasonable to assume that the residual for $x = 5.5$ would be in the pattern near the residuals for ages 4-6. That is, the residual would most likely be above the line $y = 0$, which means it would be positive. A positive residual means that the actual value of the point is above the regression line. Thus the predicted value is less than the actual value, and the prediction would be too small.

10 points for this part, scored holistically. Full credit for the correct answer along with a correct justification. 2 points max. on this question if the student's answer indicates that they don't fully understand what a residual plot is.

2. The correlation coefficient between the values assumed by two variables, *X* and *Y* is .75, the standard deviations of *X* and *Y* are 1.88 and 2.45, respectively, and the point (13, 9) is on the line of best bit.

Scoring: Each part of this question is 4 points, and in most cases it's either 4 points or 0 points, though it's conceivable that in some cases 2 points may be awarded for an incorrect answer but a correct justification (or vice-versa).

A. Find the slope of the line of best fit to the raw-score scatterplot. (5 points)

Answer

.977. $b = r\frac{s_Y}{s_X} = (.75)(2.45/1.88)$

B. Find the equation of the line of best fit. (5 points)

Answer
$y = -3.701 + .977x$. From (b), we know $y = a + .977x$. Plug in (13, 9) and find $a = -3.701$)

© Copyright 1999-2009 Apex Learning, Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning, Inc.

C. Find $\bar{y}$ given that $\bar{x} = 12$ (6 points)

Answer
8.023. The point $(\bar{x}, \bar{y})$ is on the regression line. Plug in $\bar{x}$ to get $\bar{y}$.

3. A government agency cited a manufacturing company for discrimination since 55% of males applying for a position were hired, and only 42% of the female applicants were hired. The company responded that there were two types of positions open for application, assembly line and management, and that a higher percentage of women were hired at each position: 85% to 75% for the assembly line positions and 20% to 15% for the management positions. Explain what seems to be going on here. (16 points)

Answer
This is most likely an instance of Simpson's Paradox. There has been an apparent reversal of the direction of the association between the variables when the data for the different types of jobs are combined. This can happen when data is aggregated.

16 points for this question. 8 for recognizing Simpson's Paradox, 8 for indicating a correct understanding of what can happen when you aggregate data from two two-way tables

TI-83 screens are used with the permission of the publisher. Copyright © 1996, Texas Instruments, Incorporated

© Copyright 1999-2009 Apex Learning, Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning, Inc.